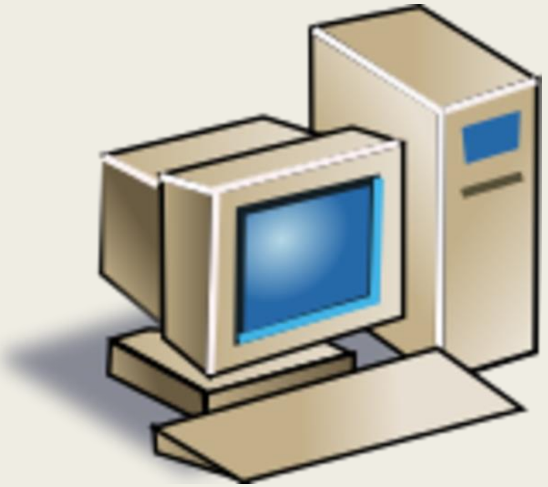
UNIVERSITEIT VAN AMSTERDAM

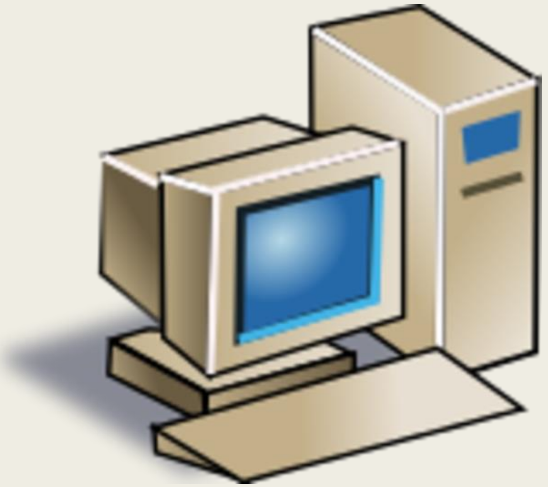
Malicious Domain Name Detection System

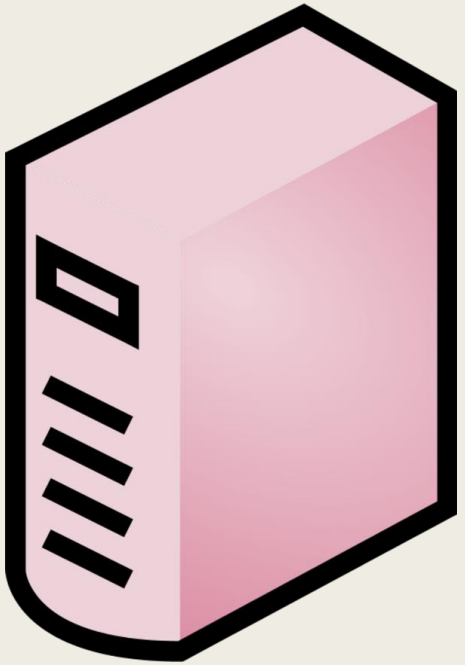
By Auke Zwaan

DNS

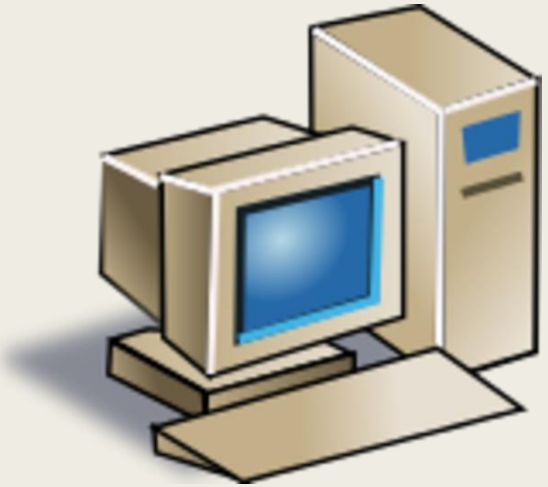


DNS

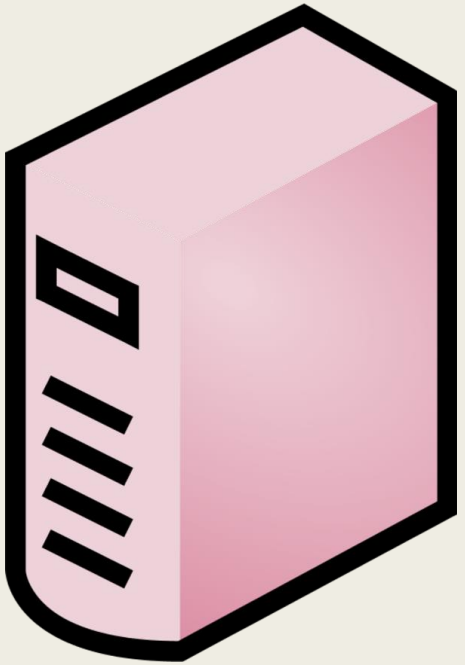




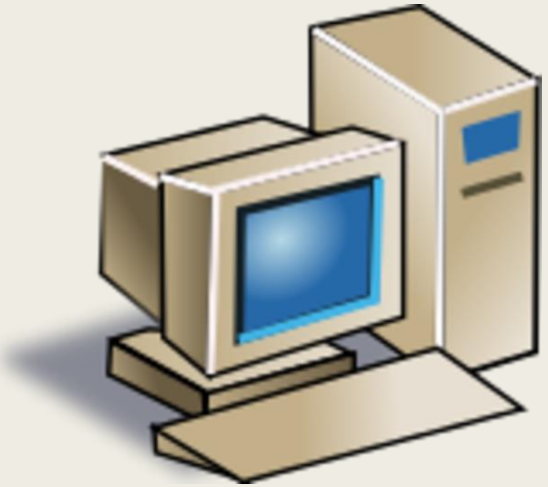
DNS



Give me **google.nl**

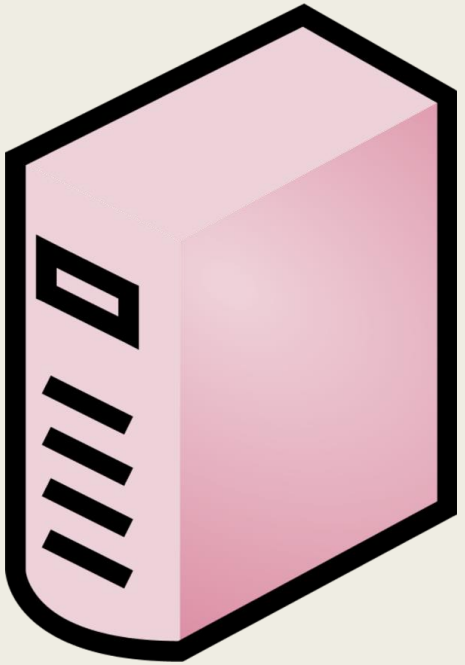


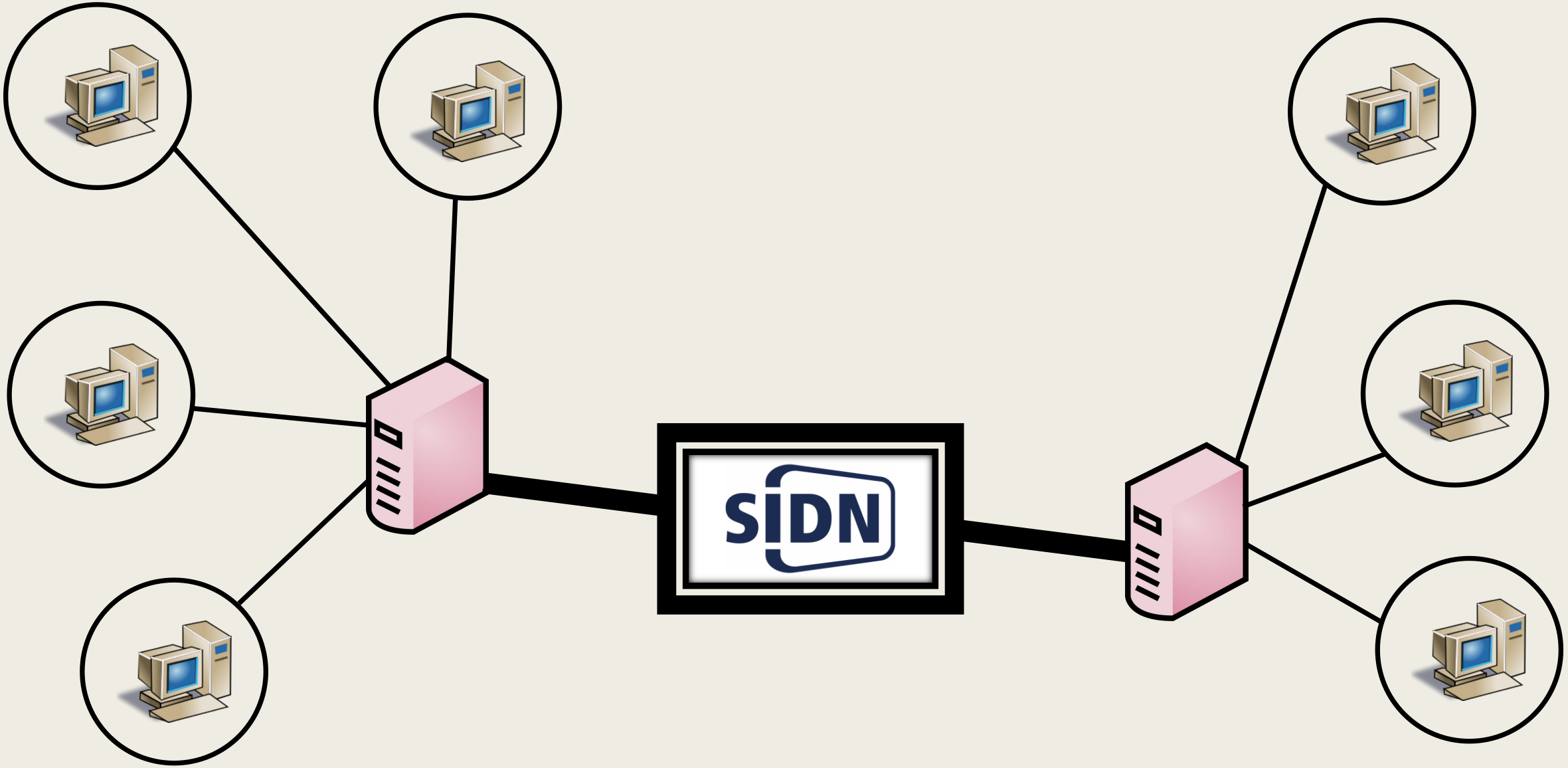
DNS



Give me **google.nl**

Okay. **64.233.166.94**




SIDN

Research Question

Is it possible to detect **malicious domains** by analyzing interrelations between **DNS resolvers** and **blacklisted domains**?

One giant DNS dataset

- All DNS requests done to *ns1.dns.nl* on January 6, 2016
- 170M+ DNS Queries (+-7GB)

DNS Data

```
1 192.168.0.105,google.nl,1452091187
2 192.168.0.106,uva.nl,1452091187
3 192.168.0.232,nu.nl,1452091187
4 145.100.104.208,os3.nl,1452091187
5 192.168.0.108,bdcrqgonzmwuehky.nl,1452091187
6 145.100.104.208,hzmksreiuojy.nl,1452091187
7 145.100.104.208,xjpakmdcfuqe.nl,1452091187
```

DNS Data

```
1 192.168.0.105,google.nl,1452091187
2 192.168.0.106,uva.nl,1452091187
3 192.168.0.232,nu.nl,1452091187
4 145.100.104.208,os3.nl,1452091187
5 192.168.0.108,bdcrqgonzmwuehky.nl,1452091187
6 145.100.104.208,hzmksreiuojy.nl,1452091187
7 145.100.104.208,xjpakmdcfuqe.nl,1452091187
```

DNS Data

```
1 192.168.0.105,google.nl,1452091187
2 192.168.0.106,uva.nl,1452091187
3 192.168.0.232,nu.nl,1452091187
4 145.100.104.208,os3.nl,1452091187
5 192.168.0.108,bdcrqgonzmwuehky.nl,1452091187
6 145.100.104.208,hzmksreiuojy.nl,1452091187
7 145.100.104.208,xjpakmdcfuqe.nl,1452091187
```

DNS Data

```
1 192.168.0.105,google.nl,1452091187
2 192.168.0.106,uva.nl,1452091187
3 192.168.0.232,nu.nl,1452091187
4 145.100.104.208,os3.nl,1452091187
5 192.168.0.108,bdcrqgonzmwuehky.nl,1452091187
6 145.100.104.208,hzmksreiuojy.nl,1452091187
7 145.100.104.208,xjpakmdcfuqe.nl,1452091187
```

DNS Data

```
1 192.168.0.105,google.nl,1452091187
2 192.168.0.106,uva.nl,1452091187
3 192.168.0.232,nu.nl,1452091187
4 145.100.104.208,os3.nl,1452091187
5 192.168.0.108,bdcrqgonzmwuehky.nl,1452091187
6 145.100.104.208,hzmksreiuojy.nl,1452091187
7 145.100.104.208,xjpakmdcfuqe.nl,1452091187
```

DNS Data

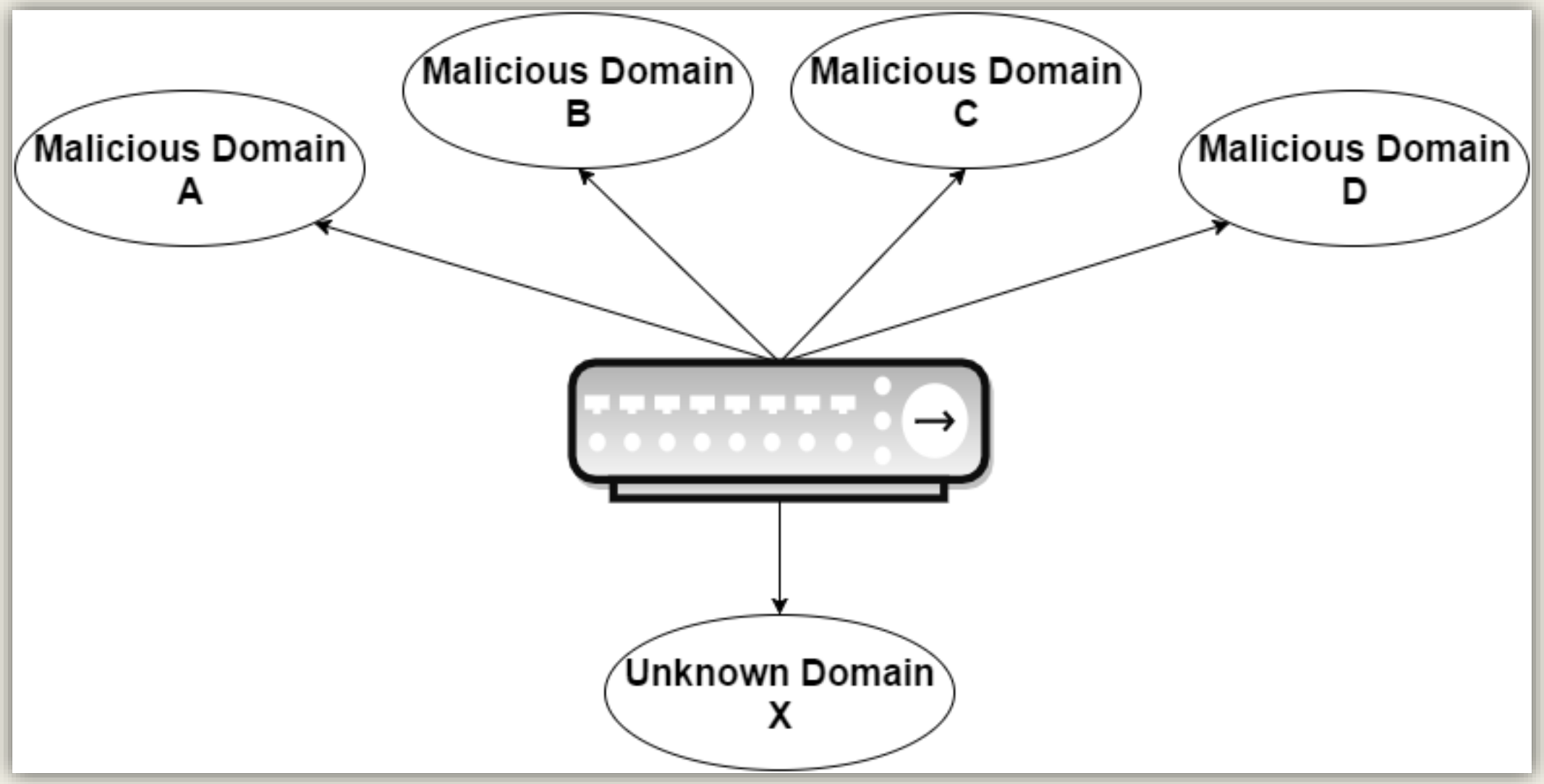
```
1 192.168.0.105,google.nl,1452091187
2 192.168.0.106,uva.nl,1452091187
3 192.168.0.232,nu.nl,1452091187
4 145.100.104.208,os3.nl,1452091187
5 192.168.0.108,bdcrqgonzmwuehky.nl,1452091187
6 145.100.104.208,hzmksreiuojy.nl,1452091187
7 145.100.104.208,xjpakmdcfuqe.nl,1452091187
```

DNS Data

```
1 192.168.0.105,google.nl,1452091187
2 192.168.0.106,uva.nl,1452091187
3 192.168.0.232,nu.nl,1452091187
4 145.100.104.208,os3.nl,1452091187
5 192.168.0.108,bdcrqgonzmwuehky.nl,1452091187
6 145.100.104.208,hzmksreiuojy.nl,1452091187
7 145.100.104.208,xjpakmdcfuqe.nl,1452091187
```

DNS Data

```
1 192.168.0.105,google.nl,1452091187
2 192.168.0.106,uva.nl,1452091187
3 192.168.0.232,nu.nl,1452091187
4 145.100.104.208,os3.nl,1452091187
5 192.168.0.108,bdcrqgonzmwuehky.nl,1452091187
6 145.100.104.208,hzmksreiuojy.nl,1452091187
7 145.100.104.208,xjpakmdcfuqe.nl,1452091187
```


Malicious Domain
A
Malicious Domain
B
Malicious Domain
C
Malicious Domain
D
Unknown Domain
X

Nice, but what is a ‘**malicious domain name**’?

Initial blacklist

- joewein.de LLC: 424 domains
- SIDN Labs Sinkhole: 15 domains
- Internet Storm Center (SANS): 14 domains
- MalwareDomainList.com: 6 domains

Total	459 domains

X

Blacklist

=

Potentially Malicious Domains

Processing the data

Source	Target	Timestamp
192.168.0.105	google.nl	1452091187
192.168.0.106	uva.nl	1452091187
192.168.0.232	nu.nl	1452091187
145.100.104.208	os3.nl	1452091187
192.168.0.108	bdcrqgonzmwuehky.nl	1452091187
145.100.104.208	hzmksreiuojy.nl	1452091187
145.100.104.208	xjpakmdcfuqe.nl	1452091187

Processing the data

Source	Target
192.168.0.105	google.nl
192.168.0.106	uva.nl
192.168.0.232	nu.nl
145.100.104.208	os3.nl
192.168.0.108	bdcrqgonzmwuehky.nl
145.100.104.208	hzmksreiuojy.nl
145.100.104.208	xjpakmdcfuqe.nl

Processing the data

Source	Target
192.168.0.105	google.nl
192.168.0.106	uva.nl
192.168.0.232	nu.nl
145.100.104.208	os3.nl
192.168.0.108	bdcrqgonzmwuehky.nl
145.100.104.208	hzmksreiuojy.nl
145.100.104.208	xjpakmdcfuqe.nl

Grouping queries, suspicious resolvers only

Source	Target
145.100.104.208	os3.nl
	hzmksreiuojy.nl
	xjpakmdcfuqe.nl
	aanrechtblad-kopen.nl
192.168.0.108	bdcrqgonzmwuehky.nl
	replicarolex.nl
	google.nl

Flagging malicious domains

Source	Target	Malicious
145.100.104.208	os3.nl	Unknown
	hzmksreiuojy.nl	Yes
	xjpakmdcfuqe.nl	Yes
	aanrechtblad-kopen.nl	Unknown
192.168.0.108	bdcrqgonzmwuehky.nl	Yes
	replicarolex.nl	Unknown
	google.nl	Unknown

Processing the data

Source	Malicious	Unknown
145.100.104.208	2	2
192.168.0.108	1	2
192.168.0.106	1	6
192.168.0.105	1	5

Defining the ***maliciousness ratio***

$$Maliciousness\ Ratio = \frac{Number\ of\ queries\ to\ \boldsymbol{malicious\ domains}}{Number\ of\ queries\ to\ \boldsymbol{unknown\ domains}}$$

Processing the data

Source	Malicious	Unknown	Ratio
145.100.104.208	2	2	1.0
192.168.0.108	1	2	0.5
192.168.0.106	1	6	0.167
192.168.0.105	1	4	0.25
192.168.0.232	4	300	0.013

Assumption 1

A **malicious resolver** is a resolver for which the *maliciousness ratio* ≥ 0.25

Processing the data

Source	Malicious	Unknown	Ratio	
145.100.104.208	2	2	1.0	
192.168.0.108	1	2	0.5	
192.168.0.106	1	6	0.167	
192.168.0.105	1	4	0.25	
192.168.0.232	4	300	0.013	

Finding malicious domains

- Get all the domains requested by **malicious resolvers**
- Filter out the domains from the initial blacklist

Assumption 2

The 100 most popular *.nl* domain names are

not malicious

Results

- **40,469** queries to malicious domains
- **8,132** suspicious resolvers, doing **85M+** queries
- **673** malicious resolvers (maliciousness ratio ≥ 0.25)
- **413** potentially malicious domains
- **392** potentially malicious domains (minus top 100)

Assumption 3

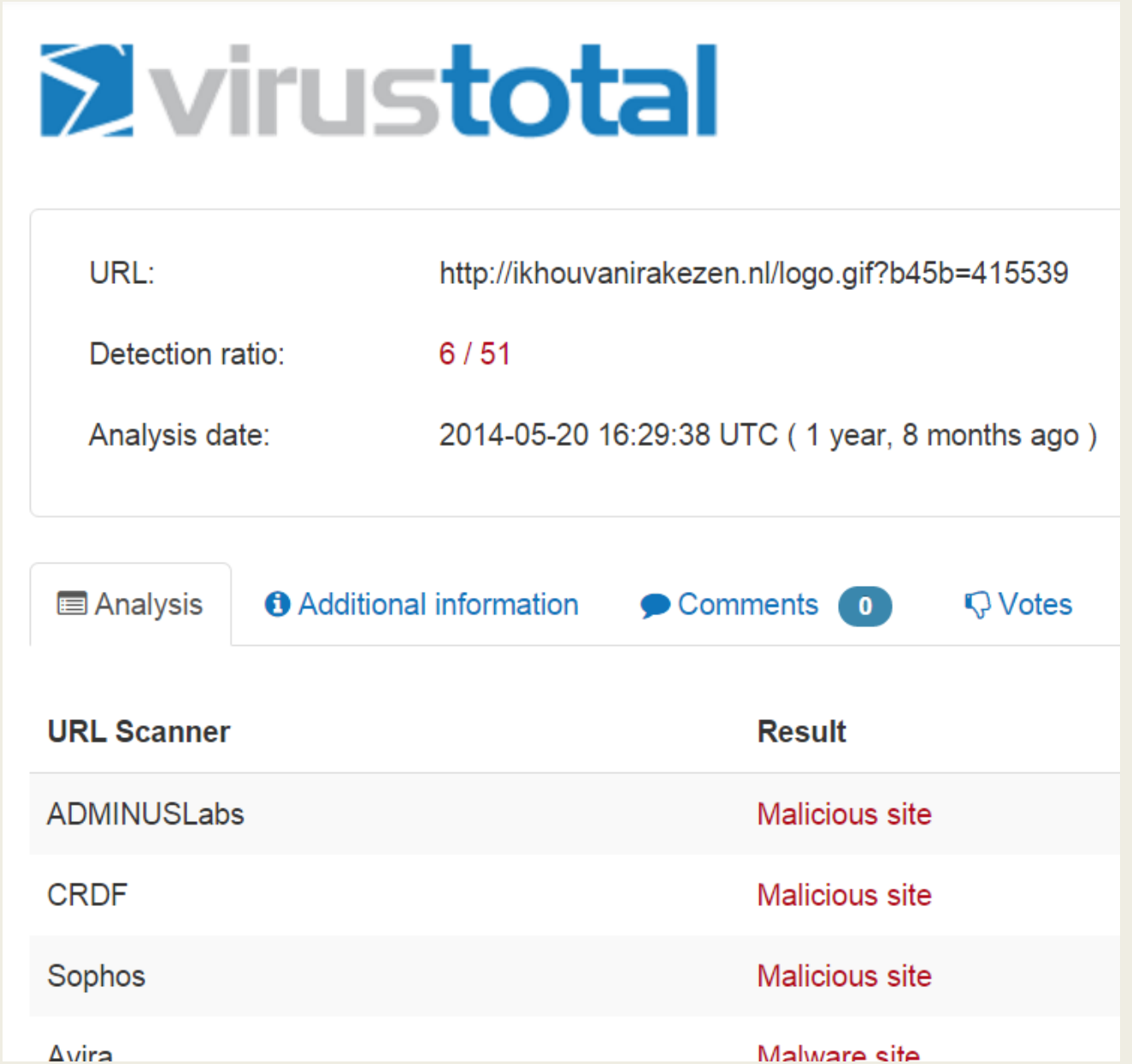
If a website has **at least one hit** in VirusTotal in the past, it is considered malicious

Example

www.ikhouvanirakezen.nl

Detected by VirusTotal

thus **true positive**



Example 2

No hits on VirusTotal

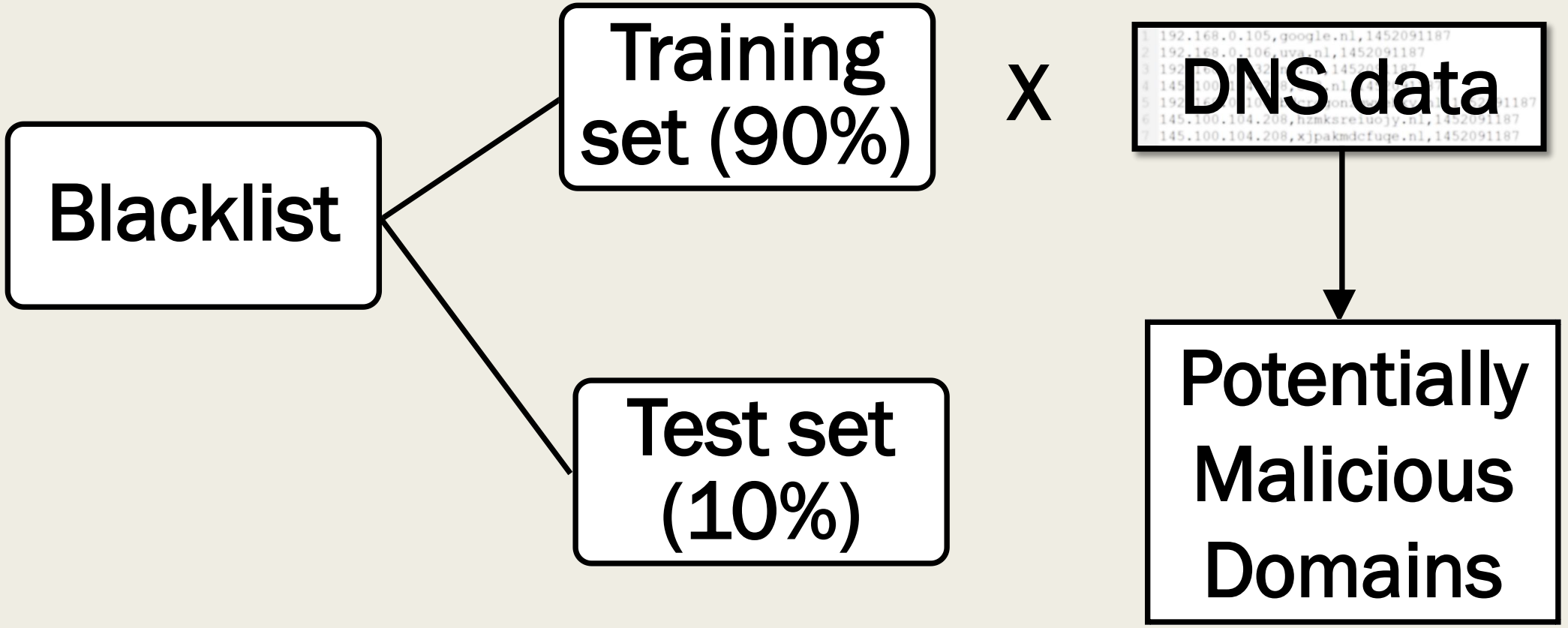
- Manual Google Search:
  - *Hits:* *Classification “Yes”*
  - *No hits:* *Classification “No”*
  - *Hosting provider:* *Classification “Possibly”*
  - *Search not feasible:* *Classifcation “Unknown”*

Malicious	Number of domains
Yes	125
No	153
Possibly	111
Unknown	3
Total	392

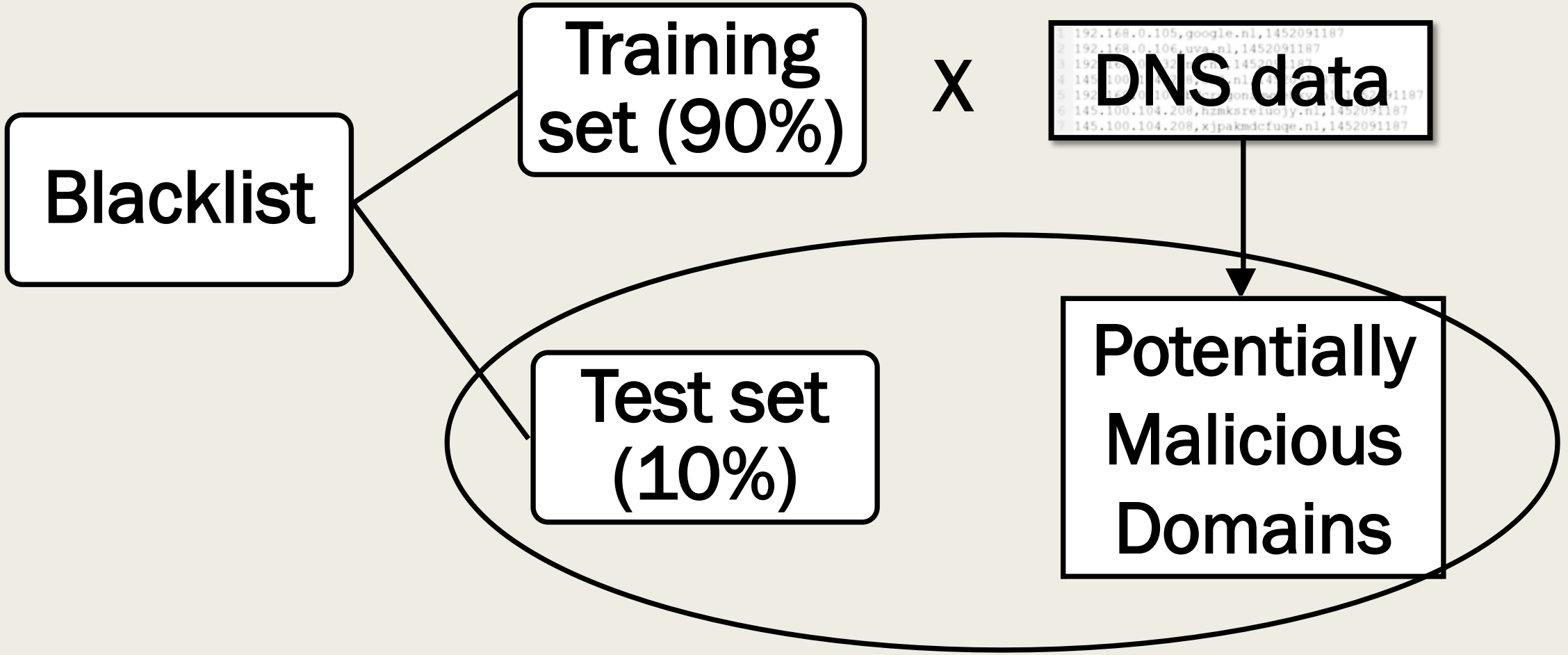
Evaluation: 32 test rounds

DNS data X Blacklist = Potentially Malicious Domains

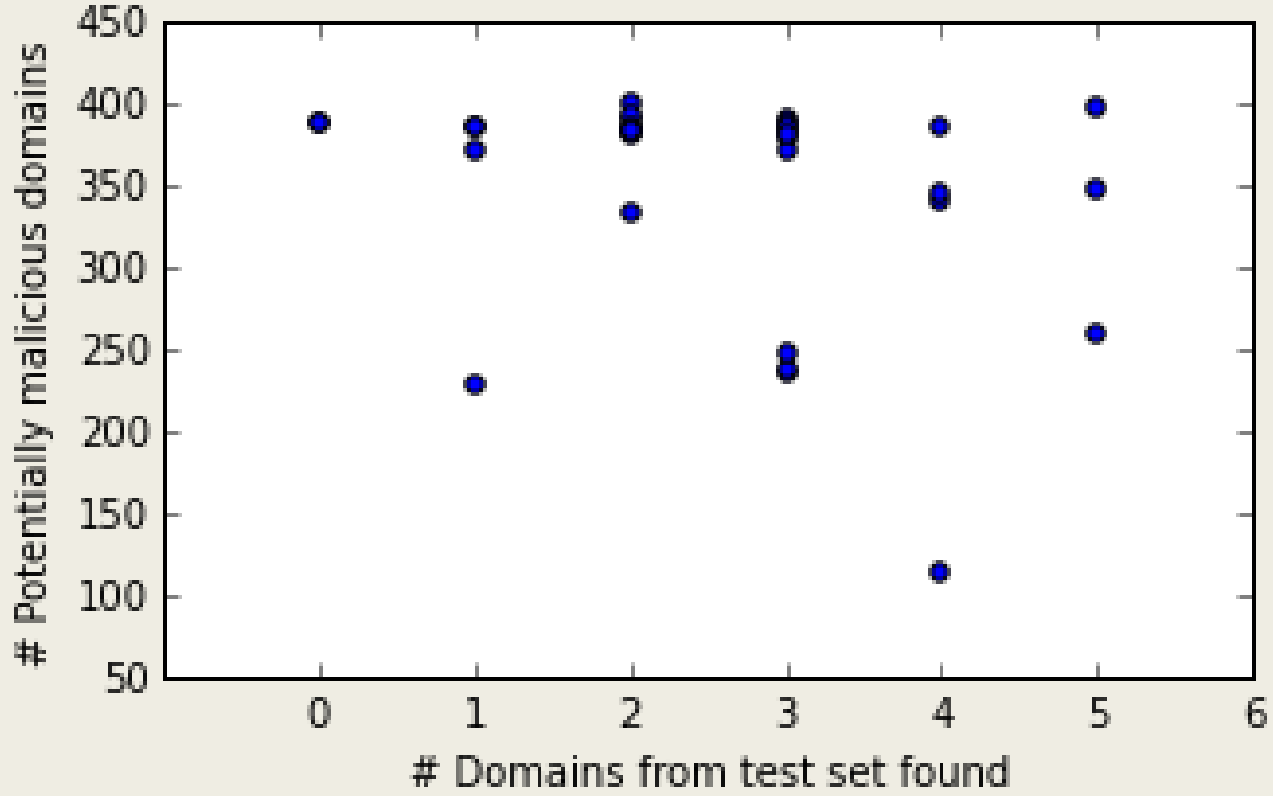
Trying to find domains from a test set



Trying to find domains from a test set



Evaluation: 32 test rounds



	Min	Max	Mean	Std
# Potentially malicious domains	114	400	349.875	68.295
# From test set found	0	5	2.594	1.316

Conclusion

- It is possible to find malicious domains by looking at spatial co-occurrence of DNS queries
- 31.8% true positives, so **not** suitable for blacklisting
- Instead, use as factor for further analysis

Future work

- Add a content analysis for each potentially malicious domain (i.e. crawling), and apply NLP
- Compare lists between different dates (datasets) and analyze commonly found domains
- Look at *whois* info for potentially malicious domains and use it for finding malicious registrars

Future work

- Extend blacklists (or run the algorithm recursively)
- Use the maliciousness ratio to identify most ‘dangerous’ resolvers
- 111x ‘Possibly’: strip out hosting providers?

Thanks for your attention!

Questions?